

Marquette University

e-Publications@Marquette

Electrical and Computer Engineering Faculty
Research and Publications

Electrical and Computer Engineering,
Department of

2-1-2016

Online Chemical Sensor Signal Processing Using Estimation Theory: Quantification of Binary Mixtures of Organic Compounds in the Presence of Linear Baseline Drift and Outliers

Karthick Sothivelr
Marquette University

Florian Bender
Marquette University, florain.bender@marquette.edu

Fabien Josse
Marquette University, fabien.josse@marquette.edu

Edwin E. Yaz
Marquette University, edwin.yaz@marquette.edu

Antonio J. Ricco
Stanford University

See next page for additional authors

Follow this and additional works at: https://epublications.marquette.edu/electric_fac



Part of the [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Sothivelr, Karthick; Bender, Florian; Josse, Fabien; Yaz, Edwin E.; Ricco, Antonio J.; and Mohler, Mary E., "Online Chemical Sensor Signal Processing Using Estimation Theory: Quantification of Binary Mixtures of Organic Compounds in the Presence of Linear Baseline Drift and Outliers" (2016). *Electrical and Computer Engineering Faculty Research and Publications*. 260.
https://epublications.marquette.edu/electric_fac/260

Authors

Karthick Sothivelr, Florian Bender, Fabien Josse, Edwin E. Yaz, Antonio J. Ricco, and Mary E. Mohler

Marquette University

e-Publications@Marquette

Electrical and Computer Engineering Faculty Research and Publications/College of Engineering

This paper is NOT THE PUBLISHED VERSION.

Access the published version at the link in the citation below.

IEEE Sensors Journal, Vol. 16, No. 3 (2016): 750-761. [DOI](#). This article is © IEEE and permission has been granted for this version to appear in [e-Publications@Marquette](#). IEEE does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from IEEE.

Online Chemical Sensor Signal Processing Using Estimation Theory: Quantification of Binary Mixtures of Organic Compounds in the Presence of Linear Baseline Drift and Outliers

Karthick Sothivelr

Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI

Florian Bender

Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI

Fabien Josse

Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI

Edwin E. Yaz

Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI

Antonio J. Ricco

Center for Integrated Systems, Stanford University, Stanford, CA

Rachel E. Mohler

Chevron Energy Technology Company, Richmond, CA

Abstract:

Compact sensor systems for on-site monitoring of groundwater for trace organic compounds in the liquid phase are currently under development in our laboratories. Potential challenges include sensor baseline drift and the presence of outliers in the data, along with difficulties extracting the contribution of individual BTEX compound (benzene, toluene, ethylbenzene, and xylenes) from the sensor response to mixtures containing multiple chemically similar compounds. As a first step, the approach presented here permits online estimation of analyte concentrations in binary mixtures of BTEX compounds in the presence of linear baseline drift and outliers. This paper investigates a sensor signal-processing approach based on estimation theory, specifically, Kalman filter (KF), extended KF, and discrete low-pass filter. The approach permits online linear baseline drift correction, filtering of outlier points, and estimation of analyte concentration(s) in binary mixtures and single analyte samples, before the sensor response reaches steady state. Sensor signals from mixtures of BTEX compounds were analyzed because these compounds are good indicators of accidental releases of fuel and oil into groundwater. Models were first developed for the sensor response so that estimation theory can be used to obtain the sensor parameters. The baseline-drift correction technique uses KF to perform online linear extrapolation or interpolation. The presented combination of sensor signal-processing techniques was simultaneously tested using actual measured data. Unknown sensor parameters and identification of analytes in samples were obtained within a relatively short period of time (8 min or less for the present sensor system), well before the sensor response reaches equilibrium.

SECTION I. Introduction

Accidental releases from fuel and oil tanks, pipelines and other sources may contaminate groundwater, lakes, rivers and oceans, potentially affecting human health [1]–[2][3][4]. Timely detection of small concentrations of hazardous chemicals in such accidental releases is of great importance for human health and has become the subject of environmental legislation. Crude oil and its refined products contain BTEX compounds (benzene, toluene, ethylbenzene and xylenes) [5] and in particular, benzene, a carcinogen, is strictly regulated [6]. Gasoline contains up to 30% of BTEX compounds [7] and is often stored in underground storage tanks (USTs) where releases can go unnoticed for long periods of time unless specific measures are taken to monitor the environment.

Currently, USTs are inspected at 2–3 year intervals by collecting groundwater samples from monitoring wells and transporting them to a laboratory for analysis [2]. This practice is time consuming and costly, prohibiting more frequent monitoring of the large majority of UST sites. Therefore, there is a need to develop compact sensor systems and signal processing methods capable of rapidly analyzing and quantifying BTEX compounds on-site. The monitoring system should be autonomous and/or remotely controlled to obviate the need for frequent site visits.

An *in-situ* chemical sensor system is currently under development in our laboratories using SH-SAW devices coated with various polymer thin films; they show promise for the detection of BTEX

compounds in trace amounts in the liquid phase [1], [8]–[9][10]. Challenges associated with the development of a compact sensor system for on-site monitoring of groundwater include sensor baseline drift, the presence of outlier points in the measured data, and the difficulties in extracting the contribution of an individual BTEX compound from the sensor response to analyte mixtures containing multiple chemically similar compounds. Sensor baseline drift is common to many chemical sensors, particularly for in-situ chemical sensors, where the sensor's environment is not controlled and environmental parameters like temperature and humidity can vary with time. These fluctuations not only cause the sensor baseline to drift, they can also induce outlier points in the sensor measurement and result in inaccurate quantification of the analytes in the mixture.

Sensor signal processing is key to the implementation of compact sensor systems. Its main purpose is to identify and quantify the target analyte(s). Key tasks for sensor signal-processing algorithms include baseline drift correction, determination of time-to-detection, extraction of steady-state information prior to attainment of equilibrium, transient information extraction, and filtering noise from the measured signal.

To correct for baseline drift, measurement of a reference signal immediately prior to exposure of the sensor to the analyte is often insufficient: if drift continues at a significant rate during analyte exposure, quantification of the analyte will be inaccurate. Several baseline-drift correction techniques allow estimation of true baseline throughout the analyte exposure, including linear extrapolation, linear interpolation, cubic interpolation, and the use of estimation theory. The use of estimation theory for sensor signal processing has already been demonstrated [11]–[12][13]. Linear extrapolation and linear interpolation are often used for sensors with rapid response times relative to the rate of drift. These two techniques rely implicitly upon the slope of the baseline remaining constant during exposure. Linear extrapolation has the advantage of requiring baseline data only prior to exposure of the sensor to the analyte. In contrast, linear interpolation requires data obtained both before and after analyte exposure, requiring longer measurement times and driving system design towards the capability for rapid flushing or purging of the sensor, but it is not surprising that linear interpolation is usually more accurate than linear extrapolation.

For sensors with longer response times, linear extrapolation and linear interpolation can still lead to poor estimates of the baseline during exposure, for example if the baseline changes drift rate or direction during exposure. In such cases, cubic interpolation and estimation theory may provide more accurate results. Estimation theory enables real-time baseline drift correction, which can drastically shorten the time required to quantify the analyte [11], [12].

Sensor signal processing is also crucial to improve time to detection. If steady-state (equilibrium) features are used, identification and quantification of analyte(s) cannot be performed until the sensor signal reaches its steady-state response. It may be undesirable to wait for steady state, especially in the presence of slowly-sorbing analytes and/or if urgent action in response to the detection of hazardous chemicals is required. Short time to detection depends on rapid extraction of steady-state information from the sensor response, particularly for liquid-phase detection where analyte mass transport is typically slower than in the gas phase.

One approach to shortening the time to detection is to use estimation theory to estimate the steady-state response well before the sensor actually reaches equilibrium with the analyte. Another approach is to use only the first few data points of the sensor response to estimate the initial derivative of the response, which is used without steady-state information to quantify the analyte [12]. This initial-derivative method, however, is prone to mass transport effects (i.e., how quickly analyte is delivered to the sensor) and may confer high noise or poor accuracy.

Typically, steady-state features are used with sensor arrays both to identify and to quantify the analyte, but valuable information contained in the sensor response transient is often overlooked. Using both transient and steady-state information can result in improved identification and increased recognition accuracy [9], [14]. This approach enables the use of the response of a single sensor device to simultaneously quantify multiple analytes in a sample, and it can also be combined with the use of a sensor array for greater certainty in component identification and/or improved quantitative accuracy.

A common approach to extract transient information is to fit the sensor response data to a single (or, for binary mixtures, dual) exponential [9] and determine the time constant(s) of the exponential fit. This approach can be combined with estimation theory to estimate response time and steady-state amplitude before the response reaches equilibrium.

Other approaches to extract transient information include transient integrals and dynamic slope [15]. Research is also directed at evaluating the feasibility of applying wavelets and wavelet-transform methods to extract sensor-response transient information [16].

To reduce or eliminate the effects of noise in measured data, a common approach, particularly for high-frequency noise in the steady-state region, is low-pass filtering. The low-pass filter preserves the low-frequency signal changes (i.e., due to exposure to the analyte) while filtering out the high-frequency noise in the data [17].

In this paper, the case of linear baseline drift is considered and estimation theory, in particular the Kalman Filter (KF), is used to compensate for such drift in sensor responses. Because the sensors studied here respond rapidly to BTEX compounds and because their baseline drift is observed to be linear on the time scale of the sensor response, a simplified baseline-drift correction method is utilized. This model uses KF to linearly extrapolate or linearly interpolate in near-real time and can be viewed as a special case of the baseline-drift correction technique presented in [11].

To filter outlier points from the measured data in real time, a combination of estimation theory (KF or Extended Kalman Filter (EKF)) and discrete low-pass filtering are used. Occasionally, outlier points are observed and must be eliminated or corrected in order to obtain more accurate results when performing the estimation process to quantify the analyte(s) present in the sample. Outlier points can be recorded if the measurement is affected by transient environmental factors, e.g. during the introduction of a new sample for measurement.

In the work reported here, estimation theory, in particular, KF or EKF (depending on the model), will be used to quantify analytes in binary mixtures of BTEX compounds in water. Although groundwater samples can contain mixtures of multiple analytes, only binary mixtures of analytes are considered. Using estimation theory, it will be shown that sensor response parameters and analyte concentrations

can be estimated in near real time, well before the sensor response reaches steady-state, thus reducing the time required for sample analysis and potentially increasing the lifetimes of sensor system components such as polymer coatings. The sensor response model of a two-analyte system will be presented and, because the two-analyte sensor response is based on a single-analyte sensor response, the single-analyte sensor response model will be briefly presented first. Based on the sensor response model of the two-analyte system, two different formulations of the state-space representation will be demonstrated and used to analyze the measured data in an actual experiment.

SECTION II. Background

A. Estimation Theory

In this paper, estimation theory and, in particular, Kalman Filter and Extended Kalman Filter are used to perform online chemical-sensor signal processing. Here we overview *KF* and *EKF*, which can be performed online as measurements are obtained [11]. The computational requirements allow these techniques to be implemented with standard microcontrollers [11], which is very important for the development of small, portable, cost-effective sensor systems that can detect target analyte(s) in near-real time.

Consider a system of the form

$$x_{k+1} = f(x_k, u_k, v_k) \quad (1a)$$

$$y_k = h(x_k, u_k, w_k), \quad (1b)$$

where x_k represents the state vector (in this work, the sensor parameters to be estimated), y_k the output vector of the system (in this work, the measured frequency shift), u_k the input vector (in this work, the unit step input, see (11b) below), v_k the process or state noise with covariance V_k , and w_k the measurement noise with covariance W_k . Assuming that the system represented by (1a) and (1b) meets the detectability criteria (i.e. if all unstable modes of the system are observable) [18], it is possible to estimate the unknown states, x_k , of the system by using only the available measured data, y_k [18], [19]. If (1a) and (1b) are linear and have the form given by

$$x_{k+1} = A_k x_k + B_k u_k + F_k v_k \quad (2a)$$

$$y_k = C_k x_k + D_k u_k + G_k w_k, \quad (2b)$$

where A_k , B_k , C_k and D_k represent the time-varying system matrices, *KF* can be used to estimate the state variables [13]. However, if (1a) and (1b) represent a nonlinear system, *EKF* has to be used to estimate the state variables. To apply *EKF*, the nonlinear system has to be linearized by performing a Taylor series expansion about the current state estimate and by neglecting the higher-order terms, leading to the following approximation:

$$x_{k+1} \cong f(\hat{x}_k, u_k, \bar{v}) + A_k e_k + F_k v_k \quad (3a)$$

$$y_k \cong h(\hat{x}_k, u_k, \bar{w}) + C_k e_k + G_k w_k, \quad (3b)$$

where e_k represents the error term (i.e. $e_k = x_k - \hat{x}_k$), \hat{x} is used to represent the state estimate, \bar{v} and \bar{w} represent the expected (mean) value of process and measurement noise, respectively, and matrices A_k , C_k , F_k , and G_k are defined as

$$A_k = \left(\frac{\partial f}{\partial x} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ v_k=\bar{v}}} \quad (4a)$$

$$C_k = \left(\frac{\partial h}{\partial x} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ w_k=\bar{w}}} \quad (4b)$$

$$F_k = \left(\frac{\partial f}{\partial v} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ v_k=\bar{v}}} \quad (4c)$$

$$G_k = \left(\frac{\partial h}{\partial w} \right)_{\substack{x=\hat{x}_k \\ u=u_k \\ w_k=\bar{w}}} \quad (4d)$$

After each measurement is made, the state estimates and error covariance are updated based on the newly acquired information using the following equations [20], [21]:

$$\hat{x}_{k+1} = f(\hat{x}_k, u_k, \bar{v}) + A_k P_k C_k^T (C_k P_k C_k^T + G_k W_k G_k^T)^{-1} \times [y_k - h(\hat{x}_k, u_k, \bar{w})] \quad (5a)$$

$$P_{k+1} = A_k P_k A_k^T + F_k V_k F_k^T - A_k P_k C_k^T \times (C_k P_k C_k^T + G_k W_k G_k^T)^{-1} C_k P_k A_k^T. \quad (5b)$$

The above formulation of *KF* and *EKF* can be used to perform online linear baseline drift correction, online filtering of outlier points, and the estimation of analyte concentration(s) in binary mixtures and single analyte samples.

B. Discrete Low-Pass Filter

A discrete first-order low-pass filter was used together with estimation theory to filter outlier points in the measured data in real time. This section contains a brief overview of the discrete low-pass filter used in the present work. The discrete first-order low-pass filter was designed by discretizing a simple continuous-time first-order low-pass filter, design details for which are given elsewhere [22]. The final form of the discrete low-pass filter used here is:

$$y_k = \left(\frac{T}{\tau + T} \right) u_k + \left(\frac{\tau}{\tau + T} \right) y_{k-1} \quad (6)$$

where τ represents the time constant of the filter, T is the sampling period, y_k represents the output at time k , y_{k-1} represents the output at time $k - 1$, and u_k represents the input at time k . Equation (6) denotes a recursive relation between the output, y_k , and the input, u_k . Therefore, by knowing the previous output value, y_{k-1} , and the current input value, u_k , the current output value, y_k , can be calculated. It should be noted that the input, u_k , represents the data point that needs to be filtered and the time constant of the filter, τ , should be set equal to the time constant of the system needing to be filtered. Thus, for the single-analyte system, the time constant of the filter, τ , should be set equal to the value of the time constant of the analyte response. For a two-analyte system, each analyte has its own time constant. The cut-off frequency of the two-analyte system should be smaller than the smallest frequency (corresponding to the largest time constant) of the analyte responses. Therefore, as an approximation, for the two-analyte system, the time constant of the filter, τ , can be set equal to

the time constant of the analyte with the largest time constant value. Note that the alternative, the process of finding the effective time constant of the two-analyte system, would be cumbersome and is therefore not utilized.

SECTION III. Theory

A. Linear Baseline Drift Correction

Linear baseline drift correction technique uses KF to perform linear extrapolation or linear interpolation online. Techniques for the former are explained first. In this approach, only data obtained prior to sensor exposure are used. Since baseline drift can be approximated as linear in the present case, it is modeled as a first-order polynomial:

$$y_baseline_k = a + bk + w_k. \quad (7)$$

where a represents the y-intercept, b is the slope of the baseline, w_k represents the measurement noise, and k represents the discrete time instant at which the baseline is measured. Using the measured data recorded before the analyte is introduced to the sensor, the constants a and b can be estimated and then used to extrapolate the baseline during the sensor response; baseline drift is corrected by subtraction. In order to estimate the constants a and b in real time using KF , the baseline drift model given in (7) was transformed into the state-space model by assigning a state variable, x_k , to the parameters a and b ; the following set of equations were obtained:

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+2}^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} \quad (8a)$$

$$y_k = \begin{bmatrix} 1 & k \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + w_k. \quad (8b)$$

Using the state-space model of the baseline drift and the measured data before analyte exposure, KF can be used to estimate the constants a and b , and the baseline can then be extrapolated during the sensor response in order to correct the measured data for baseline drift online. The baseline-corrected measurement can be found by subtracting the baseline value from the recorded measured data at a given instant in time,

$$y_corrected_k = y_measured_k - y_baseline_k. \quad (9)$$

Since linear extrapolation requires only the data obtained before analyte exposure to estimate the baseline, linear extrapolation can be performed in real time.

If several samples are measured consecutively in the course of one experiment, linear interpolation using KF can be used to estimate the baseline to provide a more accurate result. Linear interpolation using KF can be accomplished using the same state-space model given in (8). However, for linear interpolation, linear baseline drift must be estimated twice, once using the data obtained before the analyte exposure and again using the data obtained after the analyte has been removed from the sensor. The slope of the baseline, b , is then determined by taking the average of the slopes of the two baselines. For the y-intercept, a , one can assign the same value obtained for the y-intercept of the baseline estimated using the data obtained before analyte exposure. Once the constants a and b are

determined, the baseline during the sensor response can be determined and subtracted from the measurement data. Since linear interpolation requires waiting for the data obtained after the analyte has been flushed from the sensor to estimate the baseline, this approach cannot be performed in real time before the measurements are complete. However, linear interpolation can, in some cases, yield a more accurate result than linear extrapolation [11].

B. Online Filtering of Outlier Points

Outlier points in the experimental data occur due to high measurement noise, sometimes when the sensor is exposed to the analyte(s) but also due to irregular changes in boundary conditions, e.g., if the flow across the sensor is briefly stopped when switching to a new sample, or when bubble(s) are present on the surface of the device. Outlier points can be filtered in real time using a combination of a simple first-order discrete low-pass filter and *KF* (or *EKF*), the choice depending on the state-space representation of the sensor response model used (i.e., for linear state-space representation, *KF* is selected and for nonlinear state-space representation, *EKF* is selected). *KF* (or *EKF*) serves as a one-step-ahead predictor of the next measured data point. If the difference between the predicted and actual measured values exceeds a certain threshold set by the user (e.g., 0.01 kHz), the actual measured point is selected to be filtered using the discrete low-pass filter. If the difference between the predicted and actual measured values is within the threshold set by the user, the actual measured point will not be filtered using the discrete low-pass filter. Using this method, only actual outlier points in the measurement will be filtered, providing a more accurate estimate of the sensor response parameters in real time.

C. Modeling the Sensor Response

In this section, models of the sensor response to multiple analytes are presented. Since these models depend on the model of the response to single analytes, the latter is reviewed first. For each system, several assumptions were made, as outlined below. The sensor response models were then transformed into discrete-time models using Euler's first-order forward-difference equation. Moreover, based on the unknown parameters that needed to be estimated, the sensor response models were transformed into state-space models to facilitate estimation. For the single-analyte system, one state-space model was formulated; for the two-analyte system, two different state-space models were developed.

1) Single-Analyte System:

To model the single-analyte system, it is assumed that the single-analyte system obeys Henry's Law for relatively low concentrations (concentrations below 50 ppm); we showed previously that this assumption is valid [1], [8], [9]. Typically, when the sensor is exposed to a step change in the ambient concentration of an analyte, the sensor signal will change most rapidly at first and then more slowly as the system approaches equilibrium. We have shown that the process of analyte absorption is well fit by first-order model described by

$$\dot{C}(t) = -\frac{1}{\tau}C(t) + \frac{\gamma_p}{\tau}C_{amb}(t) \quad (10a)$$

$$\Delta f(t) = -aC(t), \quad (10b)$$

where $C(t)$ is the concentration of analyte in the coating at time t , $C_{amb}(t)$ is the ambient analyte concentration at time t , τ is the response time constant for a given analyte/coating combination, γ_p is the polymer-liquid partition coefficient for a given analyte, $\Delta f(t)$ is the frequency shift observed at time t , and a is the steady-state (or equilibrium) frequency shift, which is a function of the sensor platform, the sensor coating, and the analyte. Equation (10b) represents the measured frequency shift of the single analyte system at time t . Both (10a) and (10b) can be normalized by division by $\gamma_p C_{max}$ (where C_{max} represents the equilibrium ambient concentration). By defining new variables as

$$m(t) = \frac{C(t)}{\gamma_p C_{max}} \quad (11a)$$

$$u_s(t) = \frac{C_{amb}(t)}{C_{max}} \quad (11b)$$

$$\alpha = -a\gamma_p C_{max}, \quad (11c)$$

the following equations are obtained:

$$\dot{m}(t) = -\frac{1}{\tau}m(t) + \frac{1}{\tau}u_s(t) \quad (12a)$$

$$\Delta f(t) = -\alpha m(t). \quad (12b)$$

where $m(t)$ represents the normalized concentration of absorbed analyte at time t , α is the normalized steady-state frequency shift, and $u_s(t)$ represents the unit step input (for $t > 0$, $C_{amb} = 0$ and for $t > 0$, $C_{amb}(t) = C_{max}$). Equations (12a) and (12b) represent the single-analyte absorption process.

As previously indicated, most sensor systems collect data at discrete-time instants (i.e., $t = kT$, where T is the sampling period). Therefore, it is necessary to transform the continuous time model of the single-analyte sensor response given in (12a) and (12b) into a discrete-time model. Using Euler's first-order forward-difference equation, the following equations were obtained:

$$m_{k+1} = (1 - S)m_k + Su_{s,k} + v_k \quad (13a)$$

$$\Delta f_k = \alpha m_k + w_k. \quad (13b)$$

where S is the absorption rate constant (i.e., $S = \frac{T}{\tau}$). Note that in (13a) and (13b), the terms v_k and w_k are added to represent the process noise and measurement noise, respectively, which could be present in the system.

For the single-analyte system, it is assumed that the normalized concentration of the analyte, the time constant (absorption rate), and the steady-state frequency shift are the unknown parameters, based upon which equations (13a) and (13a) can be transformed into the state-space model by assigning state variables to the unknown parameters. The following set of equations represents the state-space model of the single-analyte system:

$$\begin{aligned}
x_{k+1} &= f(x_k, u_k, v_k) = \begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ x_{k+1}^{(3)} \end{bmatrix} \\
&= \begin{bmatrix} (1 - x_k^{(2)})x_k^{(1)} + x_k^{(2)}u_{s,k} + v_k \\ x_k^{(2)} \\ x_k^{(3)} \end{bmatrix} \quad (14a)
\end{aligned}$$

$$y_k = h(x_k, u_k, w_k) = x_k^{(3)}x_k^{(1)} + w_k. \quad (14b)$$

From equations (14a) and (14b), it can be seen that the state-space model describing the single-analyte sensor response is nonlinear. Therefore, to perform the estimation, EKF should be used, but the nonlinear system must first be linearized using a Taylor-series expansion. The result of linearization is given above in Eqs. (3a) and (3b), with the generalized expressions of the terms given below:

$$\begin{aligned}
f(\hat{x}_k, u_k, \bar{v}_k) &= \begin{bmatrix} (1 - \hat{x}_k^{(2)})\hat{x}_k^{(1)} + \hat{x}_k^{(2)}u_{s,k} \\ \hat{x}_k^{(2)} \\ \hat{x}_k^{(3)} \end{bmatrix} \\
h(\hat{x}_k, u_k, \bar{w}_k) &= \hat{x}_k^{(3)}\hat{x}_k^{(1)} \\
A_k &= \begin{bmatrix} (1 - \hat{x}_k^{(2)}) & -\hat{x}_k^{(1)} + u_{s,k} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
F &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
C_k &= [\hat{x}_k^{(3)} \quad 0 \quad \hat{x}_k^{(1)}] \\
G &= [1].
\end{aligned}$$

By using these generalized expressions of the terms of equations (3), the *EKF* algorithm can be applied to perform the estimation of the unknown parameters.

2) Two-Analyte System:

In order to model the two-analyte system, two main assumptions were made, the first being that the mixture obeys Henry's Law, which states that when a mixture of multiple soluble species is dilute, the sorption of any given species into the polymer does not affect the sorption of the other species in any way. Free partitioning of analytes between polymer and aqueous phase is assumed, including the implication that the sorption process is fully reversible at room temperature (i.e. only physisorption occurs). Based on our experimental observations, Henry's Law is valid for analyte concentrations below 50 ppm [1], [8], [9]. Henry's Law behavior implies that the concentration of the binary mixture in the coating at any time t is actually the sum of the concentrations of each individual analyte as they would be observed in single-analyte measurements: $C_{mixture}(t) = C_1(t) + C_2(t)$ for any time t , where the subscripts 1 and 2 represent two different analytes, and the process of analyte absorption can be assumed to be first order (similar to the single-analyte case). The second assumption is that the steady-state frequency shifts are also mutually independent, that is the frequency shift due to the mixture at

any time t is the sum of the frequency shifts due to each analyte in the mixture at that time. From these assumptions, it follows that the response times of the individual analytes in the mixture will be the same as those obtained from the single-analyte measurements. Based on these two assumptions, the process of analyte absorption of the two-analyte system can be represented by the following equations:

$$\dot{C}_i(t) = -\frac{1}{\tau_i} C_i(t) + \frac{\gamma_{p,i}}{\tau_i} C_{amb,i}(t) \quad (15a)$$

$$\Delta f(t) = -\sum_{i=1}^2 \alpha_i C_i(t), \quad (15b)$$

where all the variables are defined as above, with subscript $i = 1,2$ referring to the analytes in the mixture. Equations (15a) and (15b) were normalized and discretized as shown earlier for the single-analyte case. Equations (16a) and (16b) represent the normalized equations that describe the process of analyte absorption of a two-analyte system and equations (17a) and (17b) represent the discrete-time version of the two-analyte system:

$$\dot{m}_i(t) = -\frac{1}{\tau_i} m_i(t) + \frac{1}{\tau_i} u_i(t) \quad (16a)$$

$$\Delta f(t) = \sum_{i=1}^2 \alpha_i m_i(t) \quad (16b)$$

$$m_{i,k+1} = (1 - S_i) m_{i,k} + S_i u_{i,k} + v_k \quad (17a)$$

$$\Delta f_k = \sum_{i=1}^2 \alpha_i m_{i,k} + w_k, \quad (17b)$$

where all the variables are defined as above for the single-analyte case. Note that in (17a) and (17b) the noise terms v_k and w_k are added to represent the process and measurement noise, respectively, that are present in the two-analyte system (v_k and w_k are uncorrelated white noise with zero mean).

a) Case 1: Nonlinear Model:

As indicated earlier, two different state-space models for the two-analyte system were developed, one nonlinear and the other a linear. For the nonlinear model of the two-analyte system, the normalized concentration of each analyte and the steady-state frequency shift of each analyte are the unknown parameters that must be estimated. Note that the absorption rates (represented by time constants) of each of the analytes in the mixture does not have to be estimated because it is assumed that they are known from the single-analyte experiments. Based on the unknown parameters, equations (17a) and (17b) can be transformed into state-space representation by assigning state variables to the unknown parameters. The following set of equations represents the state-space representation of the nonlinear model of the two-analyte system:

$$\begin{aligned}
x_{k+1} &= f(x_k, u_k, v_k) = \begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ x_{k+1}^{(3)} \\ x_{k+1}^{(4)} \end{bmatrix} \\
&= \begin{bmatrix} (1 - S_1)x_k^{(1)} + S_1 u_{1,k} + v_k \\ (1 - S_2)x_k^{(2)} + S_2 u_{2,k} + v_k \\ x_k^{(3)} \\ x_k^{(4)} \end{bmatrix} \quad (18a)
\end{aligned}$$

$$y_k = h(x_k, u_k, w_k) = x_k^{(3)} x_k^{(1)} + x_k^{(4)} x_k^{(2)} + w_k. \quad (18b)$$

From (18a) and (18b), it can be seen that the state-space representation is a nonlinear model, and thus *EKF* should be used to estimate the unknown parameters. In order to apply the *EKF* algorithm, the nonlinear system has to be linearized using Taylor series expansion. The result of linearization is given in (3) with the generalized expressions of the terms given below,

$$\begin{aligned}
f(\hat{x}_k, u_k, \bar{v}_k) &= \begin{bmatrix} (1 - S_1)\hat{x}_k^{(1)} + S_1 u_{1,k} \\ (1 - S_2)\hat{x}_k^{(2)} + S_2 u_{2,k} \\ \hat{x}_k^{(3)} \\ \hat{x}_k^{(4)} \end{bmatrix} \\
h(\hat{x}_k, u_k, \bar{w}_k) &= \hat{x}_k^{(3)} \hat{x}_k^{(1)} + \hat{x}_k^{(4)} \hat{x}_k^{(2)} \\
A_k &= \begin{bmatrix} (1 - S_1) & 0 & 0 & 0 \\ 0 & (1 - S_2) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
F_k &= \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \\
C_k &= [\hat{x}_k^{(3)} \quad \hat{x}_k^{(4)} \quad \hat{x}_k^{(1)} \quad \hat{x}_k^{(2)}] \\
G_k &= [1].
\end{aligned}$$

b) Case 2: Linear Model:

For the linear model of the two-analyte system, the unknown parameters that need to be estimated are only the steady-state frequency shifts of the analytes. The normalized concentration of each analyte does not have to be estimated because by using the known time constant of each analyte from the single analyte experiment and equation (17a) with initial condition $m_{i,0} = 0$, the normalized concentration for each analyte can be approximately determined (i.e. assuming no process noise) for every discrete-time instant. As a result, a simplified linear model can be obtained and the estimation of the unknown parameters can be performed using *KF*. By assigning state variables to the unknown parameters, the following set of equations, which represents the state-space representation of the linear model of the two-analyte system, was obtained:

$$\begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+2}^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} \quad (19a)$$

$$y_k = \begin{bmatrix} m_{1,k} & m_{2,k} \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix} + [1]w_k. \quad (19b)$$

From (19a) and (19b), it can be seen that the state-space representation is a linear model, and thus for this case KF should be used to estimate the unknown parameters. It should be noted that the time-varying system matrix, C_k , depends on the values of the normalized concentration for each analyte (i.e., analyte 1 and analyte 2), which can be determined for each discrete-time instant, k , using (17a).

SECTION IV. Chemical Sensor Data Acquisition

As previously stated, the signal-processing techniques will be demonstrated on actual experimental data collected during the detection of binary mixtures of *BTEX* compounds using a guided *SH-SAW* sensor platform. The sensor data analyzed in this work were collected using a 36°-rotated *YX-LiTaO₃* guided *SH-SAW* device as the sensing platform [8]–[9][10]. This device was fabricated with 10/80-nm-thick Cr/Au multi-electrode *IDTs* designed to produce a wavelength of 40 μm , resulting in an operating frequency of 103 *MHz* for the uncoated device [10]. A dual-delay-line configuration was used to minimize the effect of temperature and other secondary effects on the sensor responses, with a metalized path between the *IDTs* to eliminate acoustoelectric interactions with the contacting ambient (e.g., due to ionic conductivity in the aqueous sample). The sensing line of each device was coated with one of three sorbent polymer coatings: poly(ethyl acrylate) (*PEA*), poly(epichlorohydrin) (*PECH*), and poly(isobutylene) (*PIB*), all purchased from Sigma-Aldrich. The polymers were deposited from solution by spin coating followed by baking for 15 minutes at 55 °C, which results in thicknesses of **1.0 μm** for *PEA*, **0.6 μm** for *PECH*, and **0.8 μm** for *PIB*. The reference line was coated with poly(methyl methacrylate) (*PMMA*) and baked for 120 minutes at 180 °C, which results in a glassy, non-sorbent coating for reference purposes. All *BTEX* analytes used in the experiment were purchased from Sigma-Aldrich and had purities of $\geq 98.5\%$.

The experimental setup used to collect the sensor data is described in [9]. The experiments were performed on both single-analyte samples and binary mixtures of *BTEX* compounds. The measurements were performed on the single-analyte samples to determine the sensitivity, σ (in Hz/ppm) and the response time constant, τ (in *s*) for each coating/analyte combination considered in this work; the average values of both σ and τ are given in [9]. The concentration of each analyte is extracted by dividing the steady-state frequency shift for that single species by the average value of σ [9].

SECTION V. Results and Discussion

Results are presented here on quantification of binary mixtures using the signal-processing techniques detailed earlier. In order to compare the performance of the two state-space models, results will be shown for two cases (nonlinear model and linear model) based on the same raw sensor data. Since the raw sensor data exhibit linear baseline drift and may sometimes contain outlier points, the linear-baseline-drift-correction and outlier-filtering techniques presented earlier were used to process and correct the raw data points before quantifying the analytes. For both cases, the process of linear

baseline-drift correction, outlier filtering, and quantifying the analytes were performed simultaneously in real time; the minimum times required to obtain a good estimate of the analyte concentrations are shown in each case.

Fig. 1 presents the raw experimental data for the response of a *SH-SAW* sensor coated with **1.0 μm PEA** to a binary mixture of 500 ppb benzene and 200 ppb ethylbenzene. The data were corrected for baseline drift before quantifying the analytes in the sample; Fig. 2(a) and Fig. 2(b) show the result obtained from the nonlinear and linear models. Table I shows the estimated concentrations acquired using just the data collected for the first 4, 5 and 6 minutes of the experiment shown in Fig. 1.

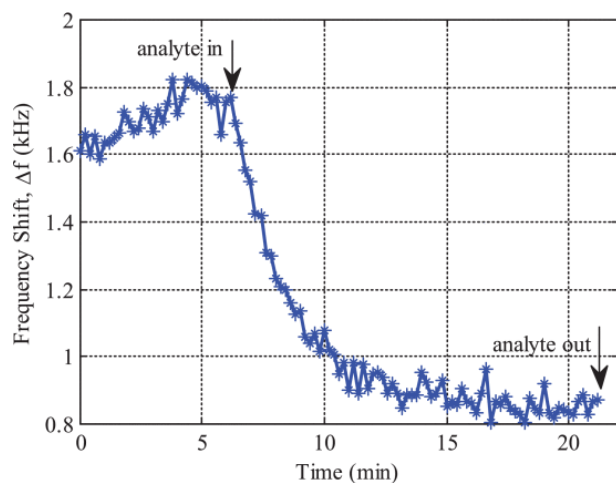
TABLE I Estimated Concentrations for a Mixture of 500 ppb Benzene and 200 ppb Ethylbenzene

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Ethylbenzene, ppb (% diff)	Benzene, ppb (% diff)	Ethylbenzene, ppb (% diff)
Initial 4 minutes	468 (6.4%)	243 (22%)	421 (16%)	228 (14%)
Initial 5 minutes	485 (3.1%)	232 (16%)	413 (17%)	231 (15%)
Initial 6 minutes	501 (0.1%)	220 (10%)	432 (14%)	224 (12%)
All data points	545 (9.0%)	191 (4.6%)	493 (1.4%)	202 (1.2%)

*SH SAW sensor coated with 1.0 μm of PEA. Both nonlinear and linear models utilized experimental data collected for the first 4, 5 or 6 minutes after the sample was introduced to the sensor, as well as all the data points (~ 12 min). Also given in the table are percentage differences between estimated and actual concentrations.

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Ethylbenzene, ppb (% diff)	Benzene, ppb (% diff)	Ethylbenzene, ppb (% diff)
Initial 4 minutes	468 (6.4%)	243 (22%)	421 (16%)	228 (14%)
Initial 5 minutes	485 (3.1%)	232 (16%)	413 (17%)	231 (15%)
Initial 6 minutes	501 (0.1%)	220 (10%)	432 (14%)	224 (12%)
All data points	545 (9.0%)	191 (4.6%)	493 (1.4%)	202 (1.2%)

*SH SAW sensor coated with 1.0 μm of PEA. Both nonlinear and linear models utilized experimental data collected for the first 4, 5 or 6 minutes after the sample was introduced to the sensor, as well as all the data points (~ 12 min). Also given in the table are percentage differences between estimated and actual concentrations.



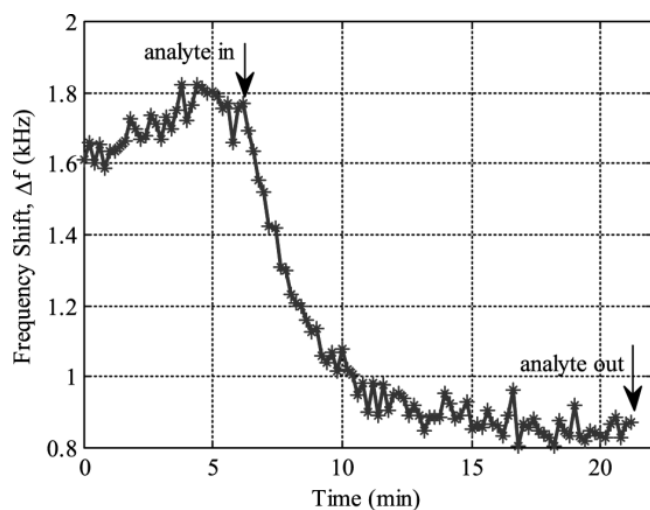
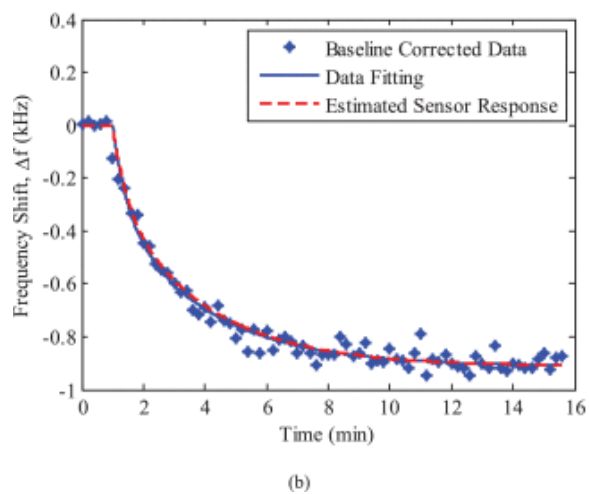
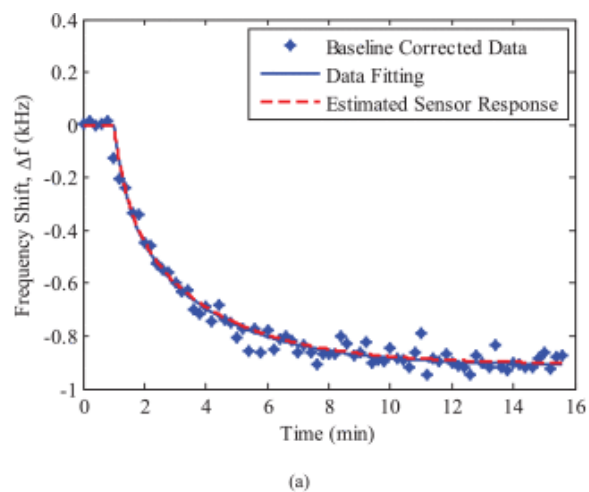
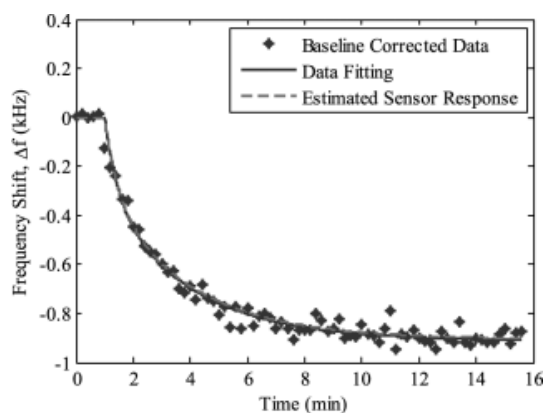
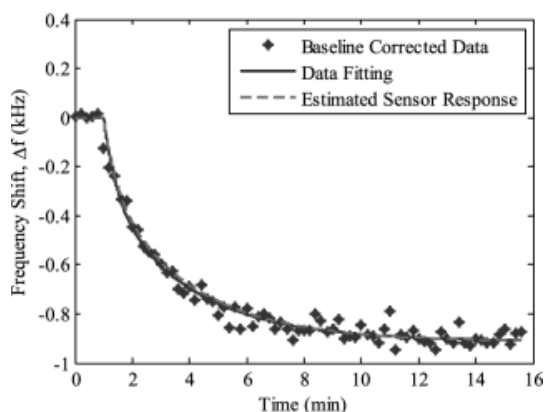


Fig. 1. Raw experimental data, including baseline drift, showing the response of a SH-SAW sensor coated with $1.0 \mu\text{m}$ PEA to a binary mixture of 500 ppb benzene and 200 ppb ethylbenzene.





(a)



(b)

Fig. 2. Baseline-drift-corrected response of a SH-SAW sensor coated with **1.0 μm PEA** to a binary mixture of 500 ppb benzene and 200 ppb ethylbenzene (blue dots), along with the estimated sensor response using the (a) nonlinear model (red curve) (estimated concentrations: benzene-545 ppb, ethylbenzene-191 ppb) and (b) linear model (red curve) (estimated concentrations: benzene-493 ppb, ethylbenzene-202 ppb). The concentrations determined by fitting the baseline corrected data using a dual-exponential fit (blue curves) are benzene-557 ppb, ethylbenzene-191 ppb.

Fig. 2(a) and Fig. 2(b) show graphically that the estimated concentrations of the analytes obtained using both models agree well with the actual concentrations. The concentrations estimated from the nonlinear model are within 9% of the actual concentrations; for the linear model, agreement is within 2%.

The results in Table I show that, using either model, the analytes can be quantified in less than half the time required for the sensor response to reach steady-state. For the nonlinear model, the estimated concentrations obtained after 5 minutes and, for the linear model, after 4 minutes, agree well (<20% difference) with the actual concentrations.

Fig. 3 shows the raw experimental data for the sensor response of a SH-SAW sensor coated with **0.6 μm PECH** to a binary mixture of 1000 ppb benzene and 500 ppb toluene (note that part of the initial drifting baseline is not shown, as indicated by the offset on the frequency axis). The results obtained using the signal-processing techniques described above are shown in Fig. 4 and Table II. Fig.

4(a) and Fig. 4(b) show the final result obtained for the nonlinear and linear models; Table II compares the estimated concentrations determined using just the data collected for the first 4, 5 and 6 minutes with those obtained using all the data.

TABLE II Estimated Concentrations for a Mixture of 1000 ppb Benzene and 500 ppb Toluene

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Toluene, ppb (% diff)	Benzene, ppb (% diff)	Toluene, ppb (% diff)
Initial 4 minutes	962 (3.8%)	425 (15%)	1008 (0.8%)	416 (17%)
Initial 5 minutes	954 (4.6%)	433 (14%)	990 (1.0%)	429 (14%)
Initial 6 minutes	959 (4.1%)	428 (15%)	984 (1.6%)	434 (13%)
All data points	939 (6.1%)	449 (10%)	973 (2.7%)	441 (12%)

*SH-SAW sensor coated with 0.6 μm PECH. Both nonlinear and linear models utilized experimental data collected for the first 4, 5 or 6 minutes after the sample was introduced to the sensor, as well as all the data points (~ 10 min). Also given in the table are percentage differences between estimated and actual concentrations.

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Toluene, ppb (% diff)	Benzene, ppb (% diff)	Toluene, ppb (% diff)
Initial 4 minutes	962 (3.8%)	425 (15%)	1008 (0.8%)	416 (17%)
Initial 5 minutes	954 (4.6%)	433 (14%)	990 (1.0%)	429 (14%)
Initial 6 minutes	959 (4.1%)	428 (15%)	984 (1.6%)	434 (13%)
All data points	939 (6.1%)	449 (10%)	973 (2.7%)	441 (12%)

*SH-SAW sensor coated with 0.6 μm PECH. Both nonlinear and linear models utilized experimental data collected for the first 4, 5 or 6 minutes after the sample was introduced to the sensor, as well as all the data points (~ 10 min). Also given in the table are percentage differences between estimated and actual concentrations.

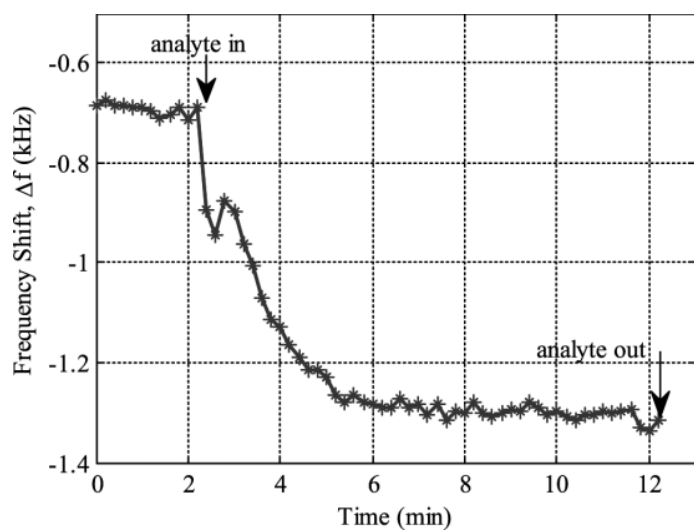
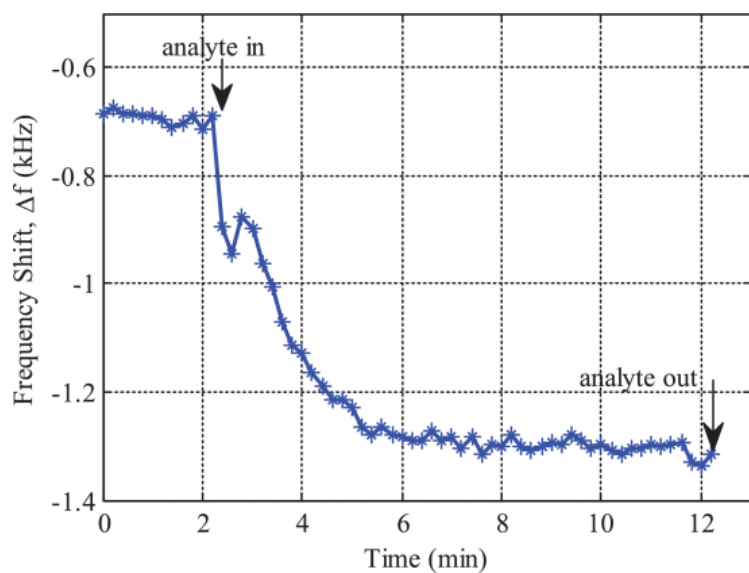
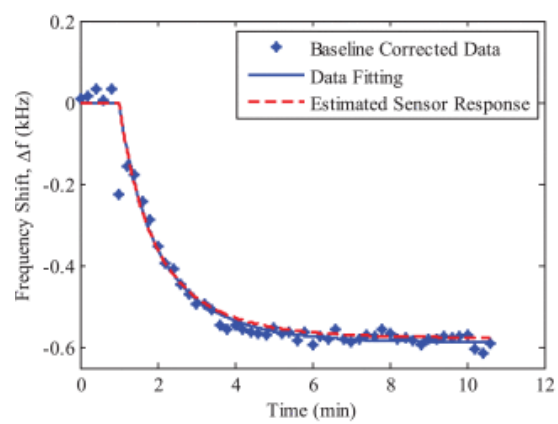
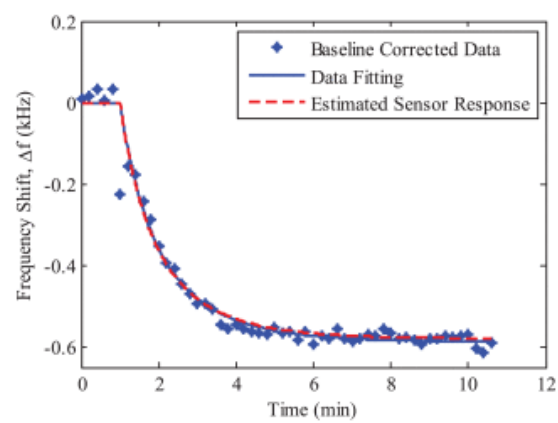


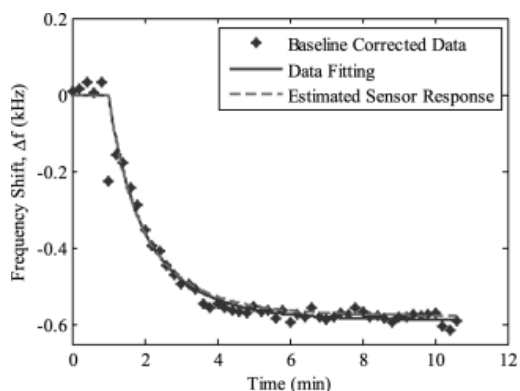
Fig. 3. Raw experimental data, including baseline drift and outlier(s), showing the response of a SH-SAW sensor coated with **0.6 μm PECH** to a binary mixture of 1000 ppb benzene and 500 ppb toluene.



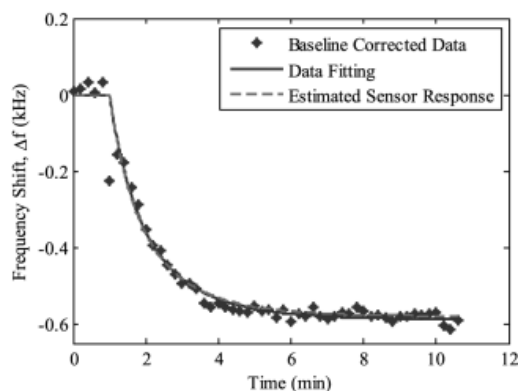
(a)



(b)



(a)



(b)

Fig. 4. Baseline-drift-corrected response of a SH-SAW sensor coated with **0.6 μm PECH** to a mixture of 1000 ppb benzene and 500 ppb toluene (blue dots) along with the estimated sensor response using the (a) nonlinear model (red curve) (estimated concentrations: benzene-939 ppb, toluene-449 ppb) and (b) linear model (red curve) (estimated concentrations: benzene-973 ppb, ethylbenzene-441 ppb). The concentrations determined by fitting the baseline corrected data using a dual-exponential fit (blue curves) are benzene-871 ppb, toluene-487 ppb.

In Fig. 4, the experimental data in Fig. 3 have been corrected for baseline drift and outliers using the techniques described above. Fig. 4(a) and Fig. 4(b) show that the estimated concentrations of the analytes obtained using both models compare well with the actual concentrations. The concentrations estimated using the nonlinear model are within 10% of the actual concentrations; for the linear model, agreement is within 12%.

The results in Table II show that, using estimation theory (either model), the analytes in the binary mixture are quantified in less than half the time required for the sensor response to reach steady-state. For both nonlinear and linear models, the estimated concentrations obtained after 4 minutes agree well (within 20%) with the actual concentrations.

Fig. 5 shows the raw experimental data for the response of a SH-SAW sensor coated with **0.8 μm PIB** to a mixture of 500 ppb benzene and 1000 ppb toluene (note that part of the initial, drifting baseline is not shown, as indicated by the offset on the frequency axis). Fig. 6(a) and Fig. 6(b) show the estimation theory results obtained using the nonlinear and linear models, respectively. Table III compares the

estimated concentrations acquired using just the data collected for the first 7, 8 and 9 minutes with those obtained using all the data.

TABLE III Estimated Concentrations for a Mixture of 500 ppb Benzene and 1000 ppb Toluene

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Toluene, ppb (% diff)	Benzene, ppb (% diff)	Toluene, ppb (% diff)
Initial 7 minutes	423 (15%)	997 (0.3%)	679 (36%)	862 (14%)
Initial 8 minutes	506 (1.1%)	961 (3.9%)	577 (15%)	892 (11%)
Initial 9 minutes	468 (6.3%)	982 (1.8%)	501 (0.2%)	915 (8.5%)
All data points	541 (8.1%)	952 (4.8%)	454 (9.2%)	929 (7.2%)

*SH-SAW sensor coated with 0.8 μm PIB. Both nonlinear and linear models utilized experimental data collected for the first 7, 8 or 9 minutes after the sample was introduced to the sensor, as well as all the data points (~ 10 min). Also given in the table are percentage differences between estimated and actual concentrations.

Portion of Response Transient used to Estimate Concentration	Nonlinear Model*		Linear Model*	
	Benzene, ppb (% diff)	Toluene, ppb (% diff)	Benzene, ppb (% diff)	Toluene, ppb (% diff)
Initial 7 minutes	423 (15%)	997 (0.3%)	679 (36%)	862 (14%)
Initial 8 minutes	506 (1.1%)	961 (3.9%)	577 (15%)	892 (11%)
Initial 9 minutes	468 (6.3%)	982 (1.8%)	501 (0.2%)	915 (8.5%)
All data points	541 (8.1%)	952 (4.8%)	454 (9.2%)	929 (7.2%)

*SH-SAW sensor coated with 0.8 μm PIB. Both nonlinear and linear models utilized experimental data collected for the first 7, 8 or 9 minutes after the sample was introduced to the sensor, as well as all the data points (~ 10 min). Also given in the table are percentage differences between estimated and actual concentrations.

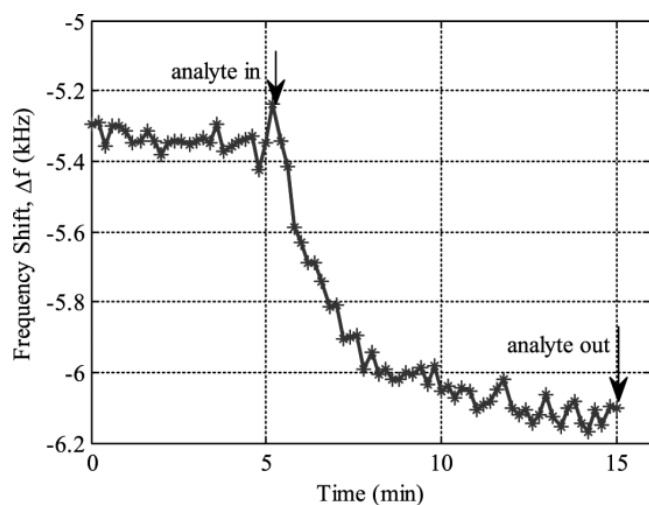
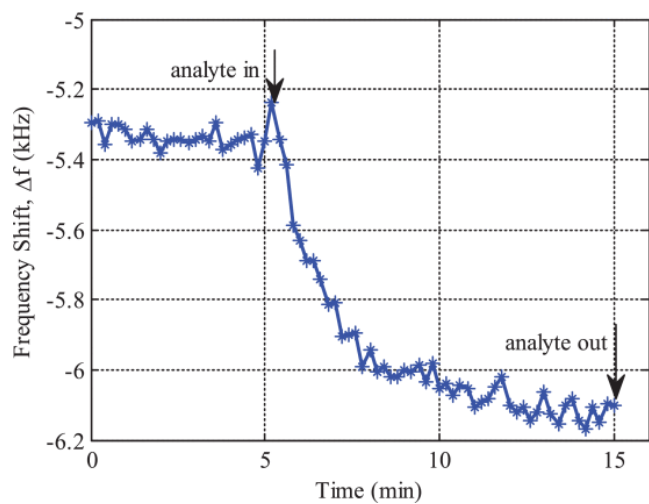
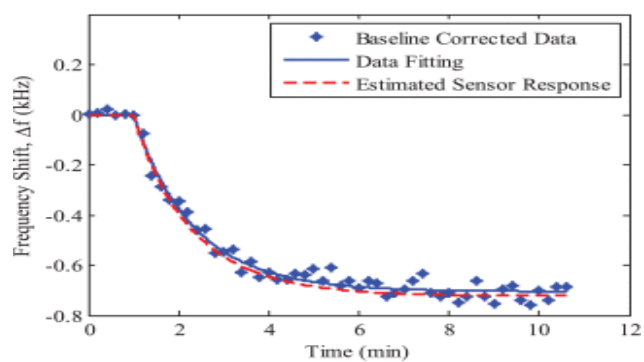
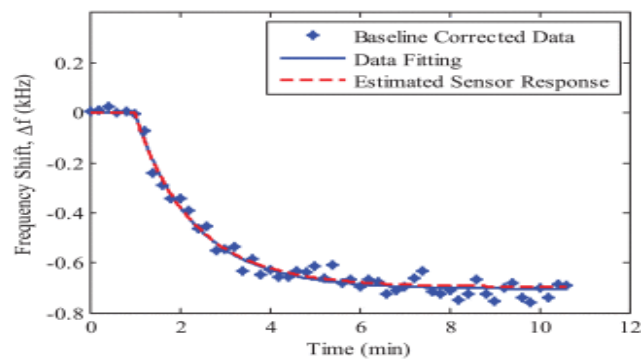


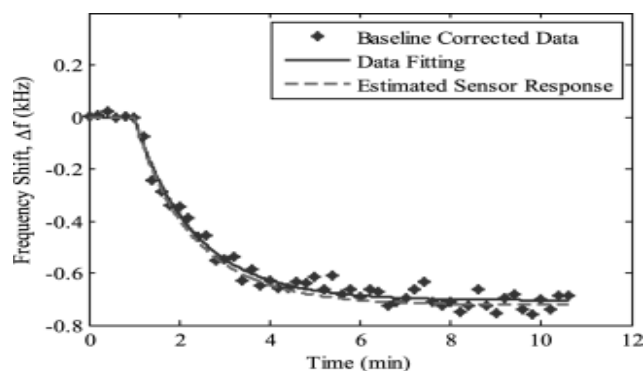
Fig. 5. Raw experimental data, including baseline drift, showing the response of a SH-SAW sensor coated with **0.8 μm PIB** to a mixture of 500 ppb benzene and 1000 ppb toluene.



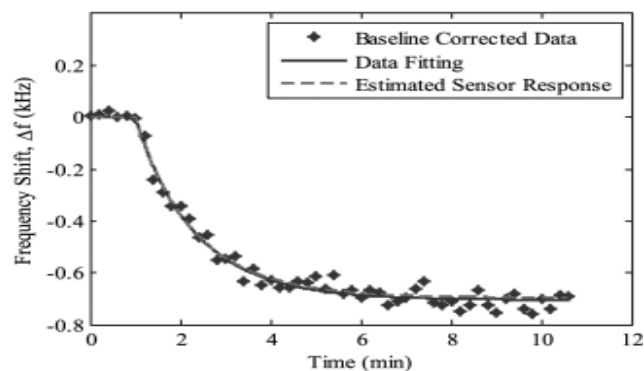
(a)



(b)



(a)



(b)

Fig. 6. Baseline-drift-corrected response of a SH-SAW sensor coated with **0.8 μm PIB** to a mixture of 500 ppb benzene and 1000 ppb toluene (blue dots) along with the estimated sensor response using the

(a) nonlinear model (red curve) (estimated concentrations: benzene-541 ppb, toluene-952 ppb) and (b) linear model (red curve) (estimated concentrations: benzene-454 ppb, toluene-929 ppb). The concentrations determined by fitting the baseline-corrected data using a dual-exponential fit (blue curves) are benzene-353 ppb, toluene-960 ppb.

In Fig. 6, the experimental data of Fig. 5 have been corrected for baseline drift and outliers using the techniques described above. Fig. 6(a) and Fig. 6(b) show that the estimated concentrations of the analytes obtained using both models compare well with the actual concentrations. The concentrations estimated using the nonlinear model are within 8% of the actual concentrations; for the linear model, the estimated and actual values are within 9%.

The results in Table III show that, using estimation theory (either model), the analytes in the binary mixture are quantified rapidly. For the nonlinear model, the estimated concentrations obtained after 7 minutes and for the linear model, the estimated concentrations obtained after 8 minutes agree well (i.e. within 20%) with the actual concentrations. Note that in Table III (as in Tables I and II), the estimated concentrations of the analytes are different for different estimation times; this is due to the presence of noise in the measured data, which affects the estimates for the concentrations obtained at different times.

Table IV summarizes the estimation results for all three polymer coatings tested. Note that all raw experimental data points were corrected for baseline drift and had outliers removed prior to being used to estimate the analyte concentrations. For each combination of model type (linear or nonlinear) and polymer film type (*PEA*, *PECH*, or *PIB*), the estimation time reported in Table IV is the one that yielded the “best” results, as indicated by the smallest sum of the squares of the percentage difference from actual concentrations. Interestingly, there appears to be a small advantage for the nonlinear model when time-to-best-result is used as a criterion (8 min on average, vs. 10.3 min average for the linear model). The number of experimental runs is too small, however, to call this advantage definitive.

TABLE IV Summary of Estimation Results* From Data Collected With Three Polymer Coatings

Polymer	Actual Concentrations of analytes, ppb	Estimated Concentrations, ppb (% difference @ time)	
		Nonlinear Model	Linear Model
1.0 μm PEA	Benzene: 500 Ethylb: 200	Benzene: 501 (0.1% @ 6 min) Ethylb.: 220 (10% @ 6 min)	Benzene: 493 (1.4% @ 12 min) Ethylb.: 202 (1.2% @ 12 min)
0.6 μm PECH	Benzene: 1000 Toluene: 500	Benzene: 939 (6.1% @ 10 min) Toluene: 449 (10% @ 10 min)	Benzene: 973 (2.7% @ 10 min) Toluene: 441 (12% @ 10 min)
0.8 μm PIB	Benzene: 500 Toluene: 1000	Benzene: 506 (1.1% @ 8 min) Toluene: 961 (3.9% @ 8 min)	Benzene: 501 (0.2% @ 9 min) Toluene: 915 (8.5% @ 9 min)

*Estimations were performed using both nonlinear and linear models. Also given in the table are the percentage differences between the estimated concentrations and actual concentrations. For each combination of film and model type, the estimation time that yielded the best results, as measured by the smallest sum of the squares of the percentage differences from actual concentrations, is reported.

Polymer	Actual Concentrations of analytes, ppb	Estimated Concentrations, ppb (% difference @ time)	
		Nonlinear Model	Linear Model
1.0 μm PEA	Benzene: 500 Ethylb: 200	Benzene: 501 (0.1% @ 6 min) Ethylb.: 220 (10% @ 6 min)	Benzene: 493 (1.4% @ 12 min) Ethylb.: 202 (1.2% @ 12 min)
0.6 μm PECH	Benzene: 1000 Toluene: 500	Benzene: 939 (6.1% @ 10 min) Toluene: 449 (10% @ 10 min)	Benzene: 973 (2.7% @ 10 min) Toluene: 441 (12% @ 10 min)
0.8 μm PIB	Benzene: 500 Toluene: 1000	Benzene: 506 (1.1% @ 8 min) Toluene: 961 (3.9% @ 8 min)	Benzene: 501 (0.2% @ 9 min) Toluene: 915 (8.5% @ 9 min)

*Estimations were performed using both nonlinear and linear models. Also given in the table are the percentage differences between the estimated concentrations and actual concentrations. For each combination of film and model type, the estimation time that yielded the best results, as measured by the smallest sum of the squares of the percentage differences from actual concentrations, is reported.

While Table IV shows the best results when minimizing the estimated concentration differences from actual values, we explored an alternative criterion: the shortest time at which both estimated concentrations come within 20% of actual values. With this approach, acceptable results are obtained within 4 min for *PEA* (linear model), 4 min for *PECH* (both models), and 7 min for *PIB* (nonlinear model). This criterion reduces the time to quantify the analytes in binary mixtures to an average of just 5 min, although there is no clear advantage for one model over the other using shortest response time as the determining factor.

SECTION VI. Summary and Conclusions

Online chemical-sensor signal-processing techniques based on estimation theory are presented. They include various steps: linear baseline drift correction, filtering of outlier points, and quantifying the

analytes in a binary mixture of organic compounds in real-time or near real-time using estimation theory. Two different models, nonlinear and linear, are developed and demonstrated for quantifying the analytes in binary mixtures.

These signal processing techniques were extensively tested on actual measured response data obtained from *SH-SAW* sensors exposed to binary mixtures of *BTEX* compounds. Based on the results obtained, these techniques do indeed accurately quantify organic compounds in binary mixtures in the presence of linear baseline drift and outliers. The estimation results obtained fall within 16% of actual concentration values in an average of 5 min, or within 10% of actual values in an average of 8 min. These results indicate that our sensor response models are accurate mathematical representations of the two-analyte system. While neither model, nonlinear or linear, is clearly superior according to the results thus far, both provide significantly more rapid and, in some cases, more accurate results than a simple dual-exponential fit.

Since these signal processing techniques involve the use of estimation theory, in particular Kalman filter (*KF*) and extended Kalman filter (*EKF*), the quantification of analytes in binary mixtures can be performed in real-time as the data are collected. The analytes in binary mixtures can be quantified in less than half the time required for the sensor response to reach steady-state with the criterion that the estimates are within 16% of the actual concentrations, or ~80 % of the time to reach estimates within 10% of the actual values.

Reducing the time to quantification of analytes allows for faster detection and more rapid response to the presence of environmental contaminants. Shortening sensor exposure time may also improve accuracy, repeatability and longevity of system components such as sensor coatings. It is important to remember that the main advantage of these signal-processing techniques is that the analytes can be quantified rapidly even in the presence of linear baseline drift and outliers. The techniques presented here can be applied to multiple analyte mixtures, provided that the assumptions made here remain valid for the mixture components and their sum, and that the response of each analyte/coating pair is known.

The demonstrated signal-processing techniques can be implemented to enable the development of a small, portable, cost-effective sensor system for field use, including in confined spaces like groundwater monitoring wells. Other potential applications include legally-required periodic groundwater monitoring near underground storage tanks, the monitoring of the plume in a sub-surface marine oil spill [3], and various spill clean-ups.

It is noted that the proposed method also enables the use of a sensor array with a smaller number of devices (2 to 3 in this case), with appropriate coatings still necessary for redundancy. Finally, it should be pointed out that the signal-processing techniques presented in this work can be used with any type of chemical sensor platform used to detect binary mixtures of analytes, and are not specific to the *SH-SAW* sensor platform. The techniques should work equally well on sensor data collected using other sensor platforms such as microcantilever-based sensors, optical chemical sensors, chemiresistors, other types of solid-state devices-based sensors, and various solid-state devices, as long as the sensor response can be modeled analytically. For the model presented in this work, the validity of Henry's law was used.

References

1. F. Bender, R. Mohler, A. J. Ricco and F. Josse, "Quantification of benzene in groundwater using SH-surface acoustic wave sensors", *Proc. 14th IMCS*, pp. 473-476, 2012.
2. *FY 2011 Annual Report on the Underground Storage Tank Program (2012)*, 2012.
3. R. N. Conmy et al., "Submersible optical sensors exposed to chemically dispersed crude oil: Wave tank simulations for improved oil spill monitoring", *Environ. Sci. Technol.*, vol. 48, no. 3, pp. 1803-1810, 2014.
4. I. M. Cozzarelli, B. A. Bekins, M. J. Baedeker, G. R. Aiken, R. P. Eganhouse and M. E. Tuccillo, "Progression of natural attenuation processes at a crude-oil spill site: I. Geochemical evolution of the plume", *J. Contam. Hydrol.*, vol. 53, no. 3, pp. 369-385, 2001.
5. J. S. Arey and P. M. Gschwend, "Estimating partition coefficients for fuel-water systems: Developing linear solvation energy relationships using linear solvent strength theory to handle mixtures", *Environ. Sci. Technol.*, vol. 39, no. 8, pp. 2702-2710, 2005.
6. *National Primary Drinking Water Regulations*, 2009.
7. J. W. Weaver, L. R. Exum and L. M. Prieto, "Gasoline composition regulations affecting LUST sites", pp. 1-37, Jan. 2010.
8. F. Bender, F. Josse and A. J. Ricco, "Influence of ambient parameters on the response of polymer-coated SH-surface acoustic wave sensors to aromatic analytes in liquid-phase detection", *Proc. Joint Conf. IEEE IFCS EFTF*, pp. 1-6, May 2011.
9. F. Bender, R. E. Mohler, A. J. Ricco and F. Josse, "Identification and quantification of aqueous aromatic hydrocarbons using SH-surface acoustic wave sensors", *Anal. Chem.*, vol. 86, no. 3, pp. 1794-1799, Jan. 2014.
10. F. Bender, F. Josse, R. E. Mohler and A. J. Ricco, "Design of SH-surface acoustic wave sensors for detection of ppb concentrations of BTEX in water", *Proc. Joint UFFC EFTF PFM Symp.*, pp. 628-631, Jul. 2013.
11. M. J. Wenzel, A. Mensah-Brown, F. Josse and E. E. Yaz, "Online drift compensation for chemical sensors using estimation theory", *IEEE Sensors J.*, vol. 11, no. 1, pp. 225-232, Jan. 2011.
12. M. J. Wenzel, "Polymer-coated and polymer-based microcantilever chemical sensors: Analysis and sensor signal processing", 2009.
13. Y.-C. Wu, Q. Chaudhari and E. Serpedin, "Clock synchronization of wireless sensor networks", *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 124-138, Jan. 2011.
14. A. Bermak, S. Brahim-Belhouari, M. Shi and D. Martinez, "Pattern recognition techniques for odor discrimination in gas sensor array" in *Encyclopedia of Sensors*, Valencia, CA, USA: American Scientific Publishers, vol. 10, pp. 1-17, 2006.
15. S. Brahim-Belhouari, A. Bermak, G. Wei and P. C. H. Chan, "On the use of the transient information for gas identification using microelectronic gas sensor", *Proc. IEEE Region 10 Conf.*, pp. 693-696, Nov. 2004.
16. K. B. Jagler, "Wavelet signal processing for transient feature extraction", Mar. 1992.
17. R. Gutierrez-Osuna, H. T. Nagle and S. S. Schiffman, "Transient response analysis of an electronic nose using multi-exponential models", *Sens. Actuators B Chem.*, vol. 61, no. 1, pp. 170-182, 1999.
18. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Mineola, NY, USA: Dover, 2005.
19. K. Reif, S. Günther, E. Yaz and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter", *IEEE Trans. Autom. Control*, vol. 44, no. 4, pp. 714-728, Apr. 1999.

20. R. E. Kalman, "A new approach to linear filtering and prediction problems", *J. Basic Eng.*, vol. 82, no. 1, pp. 35-45, 1960.
21. D. Simon, Optimal State Estimation: Kalman H_{∞} and Nonlinear Approaches , Hoboken, NJ, USA:Wiley, 2006.
22. F. Haugen, Derivation of a Discrete-Time Lowpass Filter, Trondheim,Norway:Norwegian Institute of Technology, Mar. 2008.